Supplementary Materials: Results with Different Clustering Techniques

The presentation of results centered on grouping ponderosa pine chronologies using hierarchal agglomerative clustering with Ward's method to select clusters using the Euclidean distance metric. Those particular results were selected for in-depth discussion the methodology produced groupings which were reasonable sizes (i.e. no clusters with a single chronology) and the data indicated the number of clusters within, instead of setting the number of groups prior to calculations. However, the choice of clustering technique, method for selecting clusters, and distance metric all are choices that a practitioner must make, and there is no typical "standard." In this section, the results from using other metrics and methods are presented. We begin by defining the distance metrics and giving an overview of how each method for selecting clusters occurs. Dendrograms and geographical results are subsequently provided.

Distance metrics

We considered three typical choices for distance metrics: Euclidean distance, Manhattan distance, and maximum distance. Euclidean distance is also known as the L^2 norm. It is computed by taking the sum of squared differences, element-wise, for a given pair of ring-width chronologies. Then, the square-root is taken over the entire summation. Manhattan distance is computed by taking the sum of absolute differences, element-wise, for a given pair of ring-width chronologies. The Manhattan distance is also known as the L^∞ norm. The maximum distance is found by locating the largest absolute difference, element-wise, within a given pair of ring-width chronologies and also known as the L^∞ norm.

Methods for selecting clusters

Four different methods for selecting clusters were also considered: single linkage, complete linkage, average linkage, and Ward's method. The Methods section provides an overview of Ward's method.

Single linkage, also known as nearest-neighbor clustering, combines clusters by finding the closest chronologies within each of the clusters to be combined. At the initial step, each ring width chronology constitutes its own cluster. The first step combines the chronologies which are closest together. At subsequent steps, clusters are combined by finding individual chronologies with the smallest distance measure. This method is subject to a chaining phenomena, where individual elements of the cluster may be close together but others within the same cluster may be far apart.

Complete linkage, also known as farthest neighbor clustering, combines clusters by finding the minimum of the maximum distances between individual chronologies. At the initial step, each chronology constitutes its own cluster. The first step combines the chronologies with the smallest distance metric. At subsequent steps, clusters are combined by finding the individual chronologies within each cluster which are farthest apart. Clusters are combined which have the smallest maximum distance between objects.

Average linkage combines clusters by finding the minimum average distance between clusters. The distance between two clusters is defined as the average distance between each point in one cluster to every point in the other cluster. As with the other methods outlined above, the initial step begins with each chronology constituting its own cluster.

Dendrograms

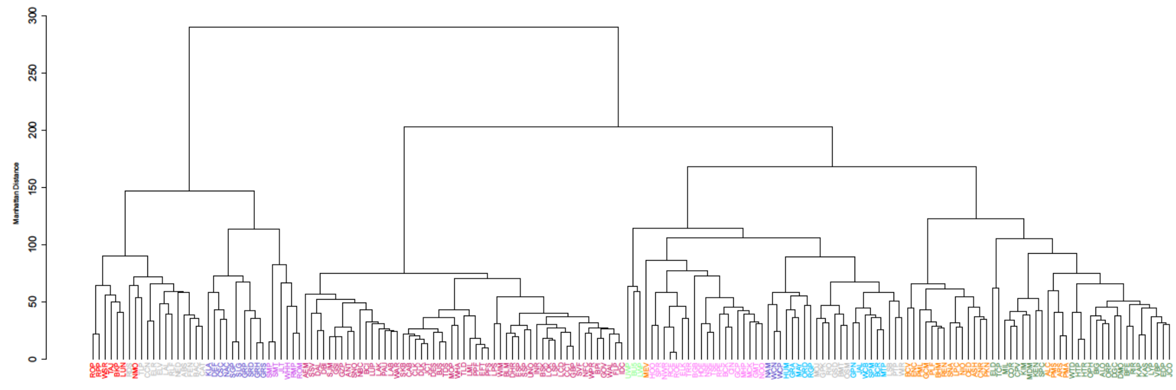

Figure S1: Hierarchal agglomerative clustering results using Ward's Method to select clusters with the Manhattan distance metric.

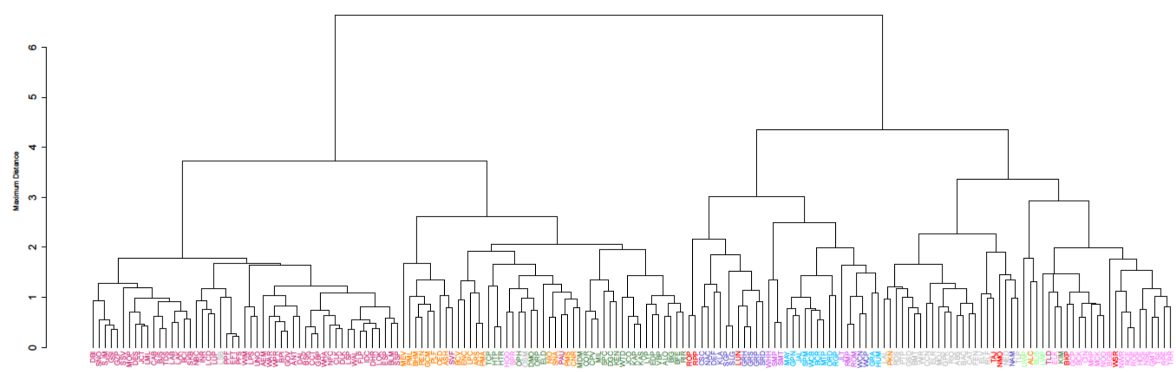

Figure S2: Hierarchal agglomerative clustering results using Ward's Method to select clusters with the maximum distance metric.

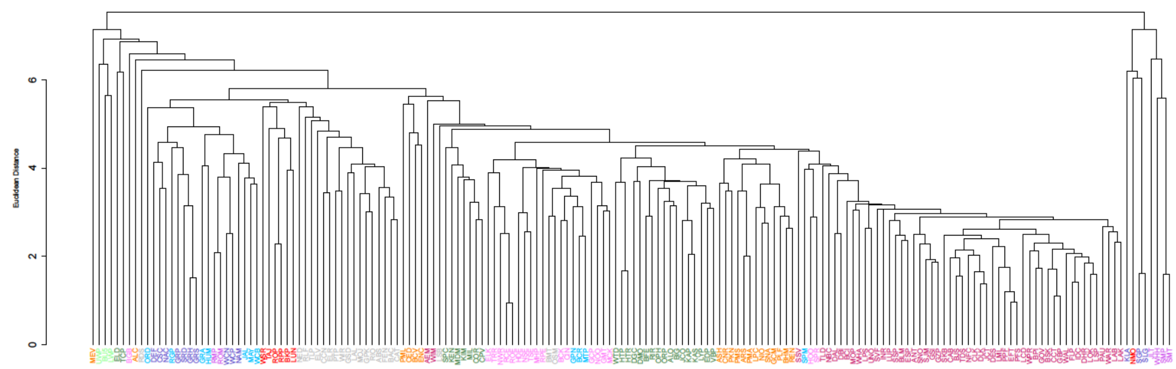

Figure S3: Hierarchal agglomerative clustering results using average linkage to select clusters with the Euclidean distance metric.

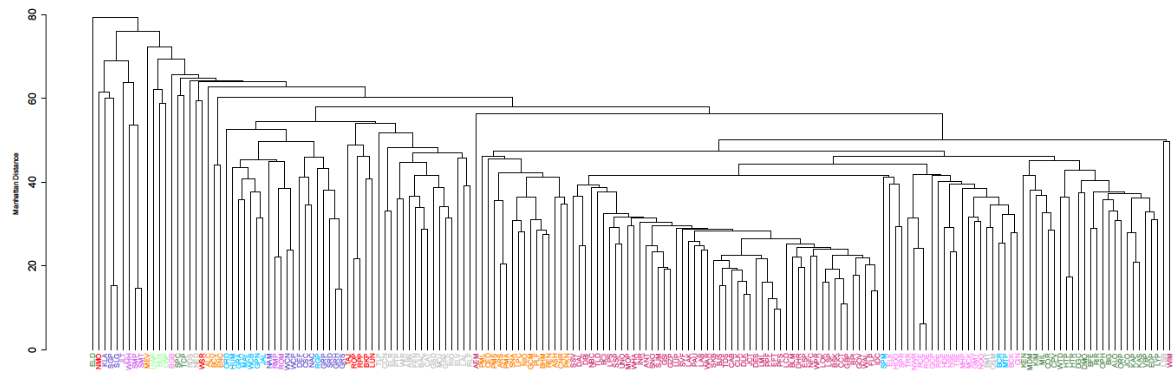

Figure S4: Hierarchal agglomerative clustering results using average linkage to select clusters with the Manhattan distance metric.

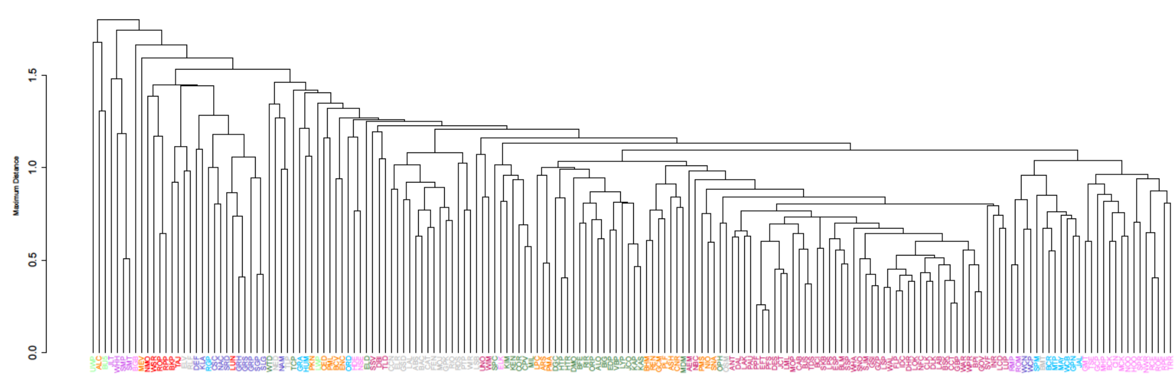

Figure S5: Hierarchal agglomerative clustering results using average linkage to select clusters with the maximum distance metric.

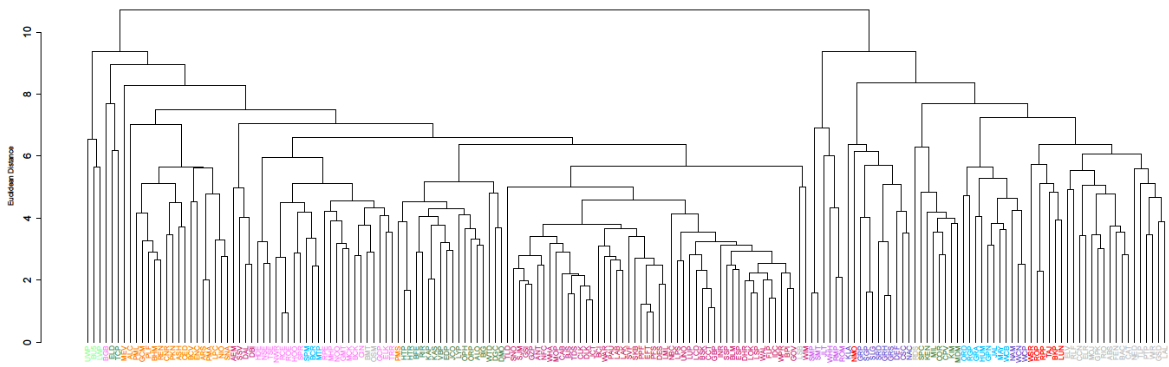

Figure S6: Hierarchal agglomerative clustering results using complete linkage to select clusters with the Euclidean distance metric.

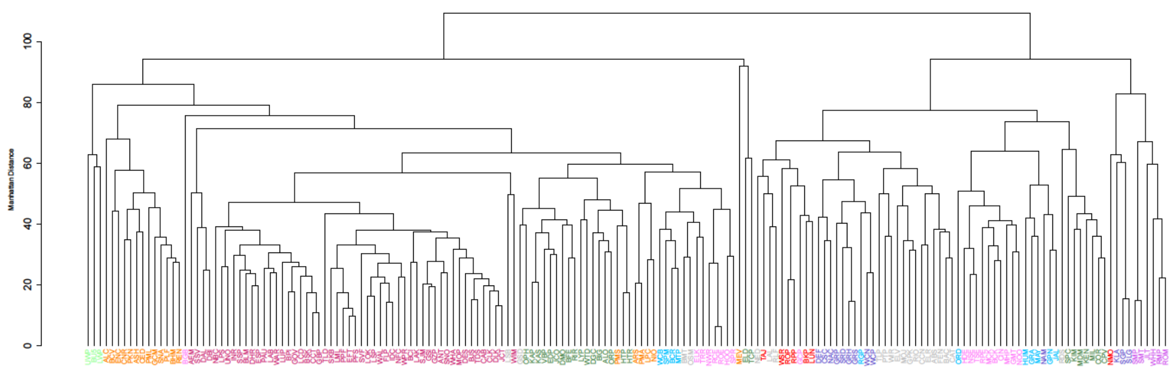

Figure S7: Hierarchal agglomerative clustering results using complete linkage to select clusters with the Manhattan distance metric.

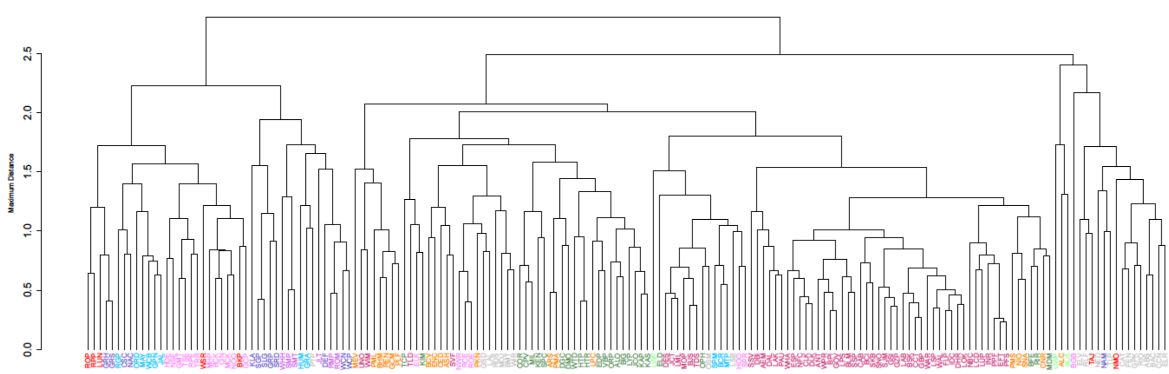

Figure S8: Hierarchal agglomerative clustering results using complete linkage to select clusters with the maximum distance metric.

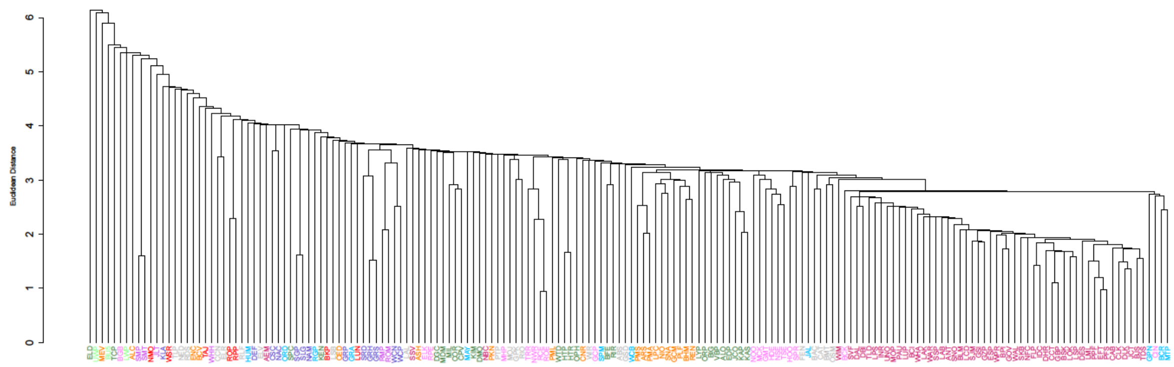

Figure S9: Hierarchal agglomerative clustering results using single linkage to select clusters with the Euclidean distance metric.

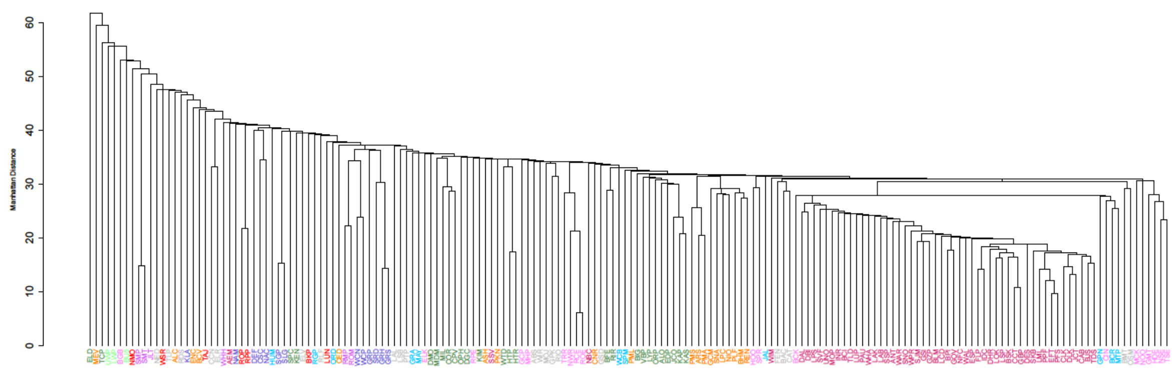

Figure S10: Hierarchal agglomerative clustering results using single linkage to select clusters with the Manhattan distance metric.

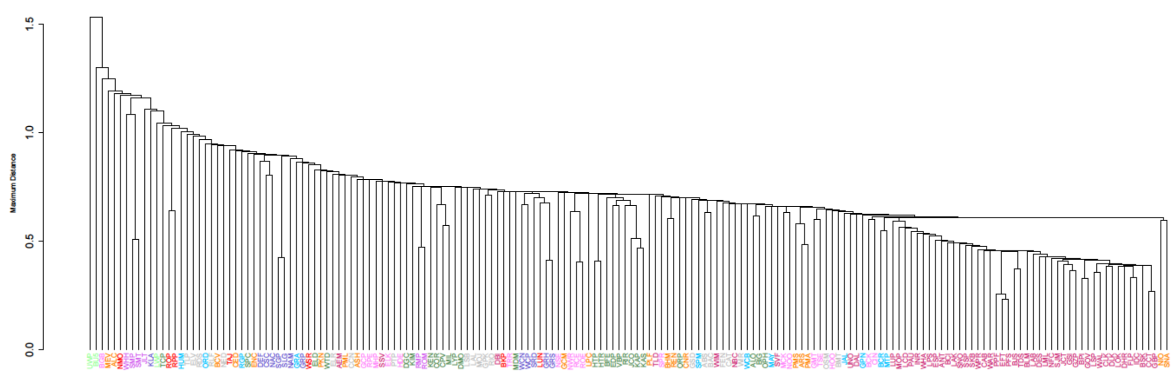

Figure S11: Hierarchal agglomerative clustering results using single linkage to select clusters with the maximum distance metric.

Qualitatively, we can see that within a method for selecting clusters, the results do not differ much for the three different distance metrics. The single linkage results all seem to exhibit the chaining phenomena, which is undesirable since many records constitute their own cluster at any given step, and at each step, most records form a single large cluster. Average and complete linkage results also have many small clusters, or clusters with single records at any given step. Because of these undesirable results, we selected Ward's method for the final analysis.

Maps

Since the dendrograms exhibited little difference due to choice of distance metric, only the results associated with Euclidean distance are provided below. Moreover, the single linkage dendrograms indicated very little utility, so those results are also excluded from this presentation. Thus, for both average and complete linkage, the bootstrap procedure was run with 10,000 bootstrap resamples to produce p-values for each branch of the dendrogram. Groupings indicated for significance levels of α=0.005 and α=0.01 are provided for each.

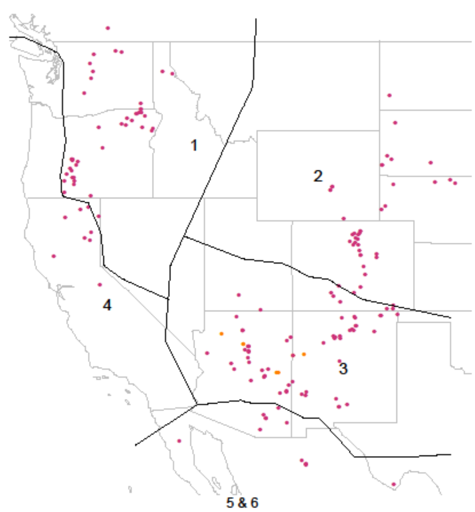

Figure S12: Hierarchal agglomerative clustering maps associated with average linkage to select clusters with the Euclidean distance metric. Two clusters are selected for both significance levels α=0.005 and α=0.01.

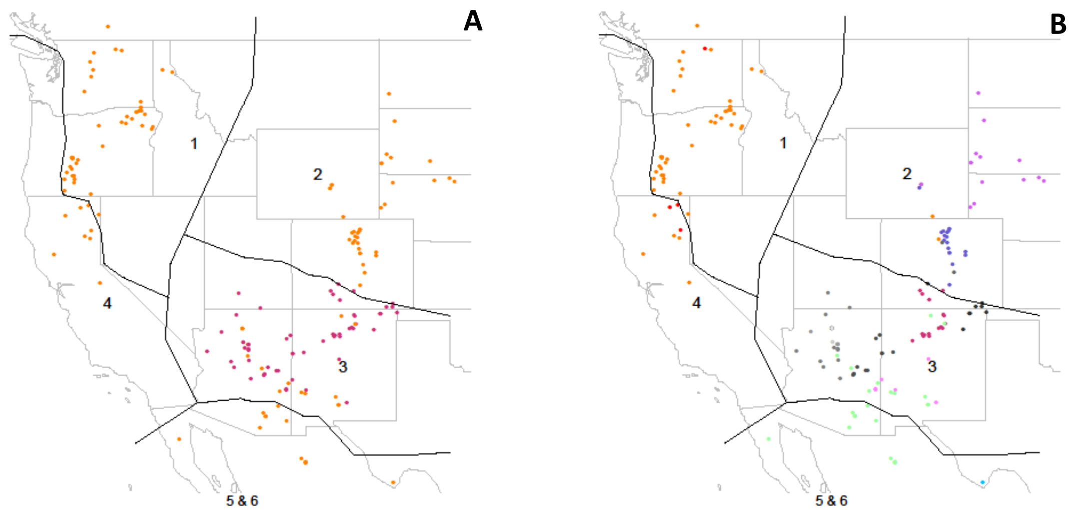

Figure S13: Hierarchal agglomerative clustering maps associated with complete linkage to select clusters with the Euclidean distance metric. A) Two clusters for α=0.005. B) Eighteen clusters for α=0.01.